Received: 3 March 2019 | Revised: 2 October 2019 | Accepted: 25 November 2019

DOI: 10.1002/cb.1804

ACADEMIC PAPER

WILEY

Agreement by design: The effect of visual harmony on responses to surveys

Ulrich R. Orth | Kristina Nickel | Rebekka Böhm | Katharina Röwe

A&F Marketing—Consumer Psychology, Kiel University, Kiel, Germany

Correspondence
Ulrich R. Orth, A&F Marketing—Consumer Psychology, Kiel University, Kiel, Germany.
Email: uorth@ae.uni-kiel.de

Abstract

The present study is one of the first to empirically examine how the visual harmony of a questionnaire can lead to measurement bias. Researchers often employ questionnaires with Likert scales to measure constructs. In this note, we examine how the design of the survey instrument, specifically, its visual harmony, can impair measurement accuracy. Two studies investigate effects of visual harmony in surveys on responses to Likert scales using paper and pencil surveys. Applying an established customer relationship management model, Study 1 employs a survey of female visitors to a grocery store (n = 115). Switching to a product and brand innovation context, Study 2 employs a survey of male and female members of a consumer panel (n = 180) to examine responses to a new e-scooter. Across studies, results indicate that assessing important consumer response constructs through visually more harmonious surveys can lead to more positive response patterns, lower scale reliability, and questionable validity, especially with females. Although these effects do not occur uniformly across measures and samples, they occur regardless of consumers' past experience with completing questionnaires, their familiarity with questionnaire design, and the naturalness and elaborateness of the visual design. Relating specific elements (e.g., text boxes, type font, shapes, and images) and relational properties of design (e.g., balance, symmetry, and coherence) to consumers' overall perception of harmony aids marketers and researchers in achieving intermediate levels to obtain realistic, reliable, and valid results.

1 | INTRODUCTION

Surveys play a fundamental role in marketing and consumer research (Singer, 2018). Numerous concepts are commonly assessed by having customers respond using paper and pencil or online surveys including Likert-type scales (Bruner, 2015). Surveys employ a number of techniques to ascertain measurement accuracy, construct validity and reliability, such as controlling the size and proximity of answer spaces (e.g., Christian, Dillman, & Smyth, 2007), number and order of options (e.g., Lee, Jones, Mineyama, & Zhang, 2002; Toepoel & Funke, 2018), the psychological distance between scale categories (e.g., Tourangeau, 2018), or the direction in which the scale runs (e.g., Nicholls, Orr, Okubo, & Loftus, 2006). One factor that is almost never controlled (Mahon-Haft & Dillman, 2010) is the visual design of the measurement instrument, specifically, its harmony. Thus, a given survey may be designed to appear visually high in harmony, may appear low in harmony, or somewhere in between. Given established effects of visual harmony on a range of viewer evaluative judgments (Haberstroh et al., 2018) and considering that harmony in visuals projects associations with agreeableness (Jiang, Gorn, Galli, & Chattopadhyay, 2016; Pittard, Ewing, & Jevons, 2007), using surveys high rather than low on visual harmony may introduce measurement bias with unknown consequences for the accuracy of results.

Harmony can be defined as "a congruent pattern or arrangement of parts that combines symmetry and balance and captures good design from a Gestalt perspective" (Henderson & Cote, 1998, 16) and as the degree to which a composition's elements form a coherent and unified pattern (Kumar & Garg, 2010). Effects of visual harmony have been studied for a variety of stimuli ranging from the simple (e.g., colors, typefonts, shapes, logos; Henderson & Cote, 1998; Henderson, Giese, &

Cote, 2004) to the more complex (e.g., packages, products, and websites; Crilly, Moultrie, & Clarkson, 2004; Orth & Malkewitz, 2008; Haberstroh et al., 2018). The findings indicate that the visual harmony of a stimulus influences a broad range of viewer cognitive and affective responses directly, as well as interactively with individual difference variables. One stream of research suggests that visual design can function as a prime, influencing consumer responses by making selected attributes focal (Mandel & Johnson, 2002). Exposure to visual harmony in a survey may therefore increase accessibility of the concept "harmony" and incorporate it in subsequent judgment and actions. Although some models suggest that making certain kinds of information accessible can invoke contrast effects (Herr, 1986), the apparent subtlety of the visual primes (Bar & Biederman, 1998) makes contrast effects in terms of viewer reactance unlikely (Mandel & Johnson, 2002).

Likert scales, perhaps one of the most widely employed scale types in consumer research (Bruner, 2015), require respondents to indicate their level of agreement. When people respond to Likert scales, an increased accessibility of harmony may increase the salience of categories on the positive side of the scale (i.e., "fully agree") relative to those on the negative side (i.e., "fully disagree"), causing a deviation. We investigate this possible effect of visual harmony in surveys on responses to Likert scales using paper and pencil surveys in two contexts (i.e., customer relationship management and new product and brand introduction). Applying established psychometric models, the questionnaires assessed consumer evaluations of a grocery store and an e-scooter using Likert scales routinely employed by researchers.

As such, our study makes at least two important contributions to the literature. First, we call attention to the visual design of surveys as a critical influencer of consumer responses. Second, we show that visual harmony in a survey's design can have a substantial impact on managerially relevant outcomes including the pattern of responses (i.e., scores of important constructs such as store and brand image, satisfaction, loyalty, product benefits, attitudes, and intentions), the reliability of scales, and the relationships among constructs.

The remainder of the paper is organized as follows. First, we discuss visual harmony as a higher order factor of design with a focus on constitutive lower order elements and critical relational properties. Next, we discuss visual priming as a conceptual background and derive hypotheses for how harmony in a survey impacts consumer responses. One field experiment and one laboratory experiment were employed to generate data for testing hypotheses. The paper concludes with a discussion of theoretical and practical implications, as well as limitations and future research opportunities.

2 | CONCEPTUAL BACKGROUND AND HYPOTHESES

2.1 | Visual harmony

According to Gestalt theory (Koffka, 1935) "what is perceived by the individual is understood by the individual as a whole or gestalt, not as component parts" (Smith-Gratto & Fisher, 1999). Every individual perceptual element has its own nature and characteristics, but the nature of individual elements alone cannot account for how a group of elements will be perceived. The pivotal point of Gestalt theory thus is that the perception of the whole pattern (or gestalt) cannot be explained from the sum of its parts. This is often stated as "the whole is greater than the sum of its parts" principle (Chang & Nesbitt, 2006). According to this principle, consumers group visual elements that are close together or that look or feel as if they belong together to derive a larger meaning (Kumar & Noble, 2016). This perceptual tendency also allows people to identify meaningful wholes (Hekkert, 2006). To explain how viewers organize individual elements into groups, perceiving and recognizing patterns, Gestalt scholars distinguish between lower level elements and higher order factors of design (e.g., Henderson & Cote, 1998; Kumar & Garg, 2010).

Important to our work, the higher order factor of harmony is not merely the sum of its parts (elements) but rather a holistic configuration that appears coherent, unified, and congruent (Lauer, 1979). As such, visual harmony is a pivotal but one higher order design factor; other important factors include naturalness and elaborateness (Henderson et al., 2004).

Viewer processing of visual design harmony follows a sequence, an early stage where simple elements such as color and shape are discovered and delineated, an intermediate stage where elements are grouped together to form a relational unit such as coherence, balance, and symmetry and a late stage where objects and configurations are associated with meaning (Chatterjee, 2010). In line with this sequential processing, one of the main functions of human vision is to group and organize objects, yielding perceived levels of harmony (Ramachandran & Hirstein, 1999). Functional magnetic resonance imaging studies suggest that successfully forming connections between the various elements may be inherently rewarding, providing "a pleasant 'aha' sensation" (Ramachandran & Hirstein, 1999).

Consistent with this conceptualizing, designers compose visuals by combining basic visual elements, such as shapes, materials, and colors, which consumers decode by aggregating elements into more complex higher order design factors (Noble & Kumar, 2010). In creating surveys, researchers select, combine, and organize visual elements such as text boxes, shapes, background pictures, typefaces, and colors as well as relational properties to create higher order aggregate impressions of design. Different than previous studies on aesthetics (Mahon-Haft & Dillman, 2010), background pictures and colors (Mandel & Johnson, 2002), or the visual presentation of answer categories (Christian et al., 2007), our research focuses on the higher order factor of visual harmony.

Outside the survey realm, empirical evidence from services marketing, retailing, and other fields emphasizes visual harmony as a key influencer of viewer responses for such diverse stimuli as logos (Henderson & Cote, 1998; Van der Lans et al., 2009), typefaces (Henderson et al., 2004), packages (Haberstroh et al., 2018; Orth & Malkewitz, 2008), and products (Kumar & Garg, 2010). With surveys, visual design elements contributing to overall impressions of harmony include colors and contrasts (Schloss & Palmer, 2011), typefaces (Henderson et al., 2004), spacing (Christian et al., 2007), borders

(Tourangeau, Couper, & Conrad, 2004), and overall layout (Orth & Malkewitz, 2008). Taken together, extant research substantiates that consumers perceive visual stimuli, including surveys, by aggregating details into a higher order composition of visual harmony, thereby impacting their responses.

2.2 | Visual harmony as a prime

We expect that visual harmony in a survey functions as a prime to impact consumer responses. Although the term "priming" is used to describe several distinct phenomena in marketing and psychology (i.e., semantic, categorical, and feature priming), they all share the same underlying mechanism. Specifically, exposure to some prior stimulus, the prime, increases the accessibility of information already existing in memory. Numerous studies have verified this increase in accessibility (see Minton, Cornwell, & Kahle, 2017 for a review). In feature priming, which we use in our study, a subject is exposed to a prime that highlights a particular feature, and this feature is then weighted more heavily in evaluation (Mandel & Johnson, 2002). The capability of visual harmony to activate associated concepts has been verified with marketing stimuli, where the level of harmony present in a visual primed interpersonal harmony and agreement (Haberstroh et al., 2018). This increased accessibility of harmony, and agreement should similarly affect consumer response to surveys relating to these concepts.

2.3 | Likert scales and effects of harmony

In measuring consumer response, many constructs (see Bruner, 2015) are commonly assessed by having participants respond using linear Likert (1932) scales. A fundamental feature of Likert scales involves respondents indicating their level of agreement. When people respond to Likert scales, an increased accessibility of harmony may thus increase the salience of categories on the positive side of the scale (i.e., "fully agree") relative to those on the negative side (i.e., "fully disagree"), causing a deviation. Specifically, we expect three kinds of differences in harmony to be reflected in response to Likert scales.

First, different patterns of responses to Likert items might occur in surveys high versus low in visual harmony. Given that Likert scales require respondents to indicate their level of agreement, visual harmony in a survey may lead to an increased salience of categories on the positive side of the scale (i.e., "fully agree") relative to those on the negative side (i.e., "fully disagree"), hereby leading consumers to report more positive answers for the constructs assessed. Therefore:

Hypothesis 1 *Consumers will exhibit more positive scores for constructs assessed on Likert scales when the survey is high (rather than low) in visual harmony.*

Second, scores from Likert scales might be less reliable when surveys are more rather than less harmonious in design. An increased salience of agreement (as primed by visual harmony) should not only lead consumers to report more positive answers, but should additionally result in a tighter agglomeration of answers around those positive scores. In other words, higher levels of agreement should come with smaller deviations from those scores, suggesting smaller variances in scale items. In turn, interitem and item-to-total correlations should be smaller, resulting in lower internal consistency reliability of multi-item scales. We expect:

Hypothesis 2 *Reliability will be smaller for multi-item (Likert) scales when the survey is high (rather than low) in visual harmony.*

Third, theoretically posited relationships among variables measured by Likert scales might show different patterns of association across more versus less harmonious survey designs. In other words, the construct validity of a Likert-measured variable might be restricted to specific harmony levels of visual design. Construct validity involves the network of associations between a construct and other constructs predicted by theory. Similar to the effect of harmony on scale reliability, smaller variances in construct measures (as a result of greater agreement primed by visual harmony) might be detrimental to the associations between constructs. Therefore:

Hypothesis 3 *Construct validity will be smaller for concepts assessed through Likert scales when the survey is high (rather than low) in visual harmony.*

Beyond those three kinds of differences, additional differences may exist in how males and females perceive and respond to visual harmony (e.g., Aspara & Van Den Bergh, 2014). Regarding perception, "male" designs are often associated with angular shapes and low harmony, whereas "female" designs typically exhibit more rounded shapes and higher harmony (Moss, 2009). Regarding response, females prefer visually harmonious paintings, whereas males tend to favor low harmony in paintings (Chamorro-Premuzic, Burke, Hsu, & Swami, 2010). Similarly, females respond more positively to harmonious designs and are more likely to note small details and elements of harmony (Xue & Yen, 2007). Females also prefer low contrast between colors, matching color combinations, and little variation in colors, all characteristics of harmonious designs (Orth & Malkewitz, 2008). Conceptually, gender differences in viewer response to visual harmony tie in with the selectivity hypothesis (see Meyers-Levy & Loken, 2015, for a detailed review). Specifically, differences exist between females and males in the comprehensiveness of stimulus processing: Females tend to process visual information more comprehensively, aggregating single elements into a fuller picture by elaborating on their interrelationships (Meyers-Levy, 1989). In contrast, males are more selective processors, relying on specific elements (Darley & Smith, 1995). Given the pivotal role of relational properties inherent to visual harmony, female consumers should thus have a superior ability to detect and then respond to harmony. Therefore:

Hypothesis 4 *Consumer gender will interact with a questionnaire's visual harmony to influence (a) response pattern, (b) reliabilities, and (c) construct validity such that effects will be more pronounced with females than with males.*

3 | EMPIRICAL STUDIES

3.1 | Pilot study

A pilot study (N = 30) aided in creating two versions of a questionnaire, one high and the other low in visual harmony. Elements modified to manipulate visual harmony included colors and contrasts (Schloss & Palmer, 2011), typefont (Henderson et al., 2004), spacing (Christian et al., 2007), borders (Tourangeau et al., 2004), and overall layout (Orth & Malkewitz, 2008). Analysis of variance yielded a significant effect of the questionnaire design on a two-item measure of visual harmony (F[1,29] = 148.47, p < .001), with the version designed for high harmony scoring higher than the version designed for low harmony (M = 3.44, SD = 1.46 vs. M = 6.12, SD = 0.78). Appendix A illustrates the treatments.

3.2 | Study 1

In the main study, 120 visitors to a grocery store (mean age of 44.7 years) were randomly intercepted during different days of the week and completed the questionnaire in exchange for a small incentive. As an initial test to our hypotheses, only female consumers were recruited. Five participants who gave a "not applicable" response to any question, provided incomplete data or exhibited unreasonable response pattern (all checkmarks consistently on the extreme left or extreme right side of the scale) were excluded from the analyses, leaving 115 data sets for subsequent analyses. Fifty-eight participants completed the survey in a high harmony design, and 57 participants completed a low harmony design survey (see Figure A1).

The survey employed previously developed and validated multi-item measures of store image (Chowdhury, Reardon, & Srivastava, 1998), customer satisfaction (Sirdeshmukh, Singh, & Sabol, 2002), and loyalty (Zeithaml, Berry, & Parasuraman, 1996) as well as a single item intended to capture a person's expertise in completing questionnaires. All measures were 7-item Likert scales with a left–right descending order, from 7 (fully agree) to 1 (fully disagree). Agreeing with an item thus indicated more positive image, greater satisfaction, higher loyalty, and greater expertise.

3.2.1 | Pattern of responses

To test the visual harmony effect including its robustness, we conducted analyses of covariance with consumer expertise in completing questionnaires included as a covariate. There was no effect of expertise (F[1,113] = .17; p = .682). Further results revealed that image was better for the high harmony questionnaire than for the low harmony questionnaire in terms of convenience (F[1,113] = 10.14; p = .002; η^2 = .08; $M_{\text{low harmony}}$ = 5.62, SD = 1.18 vs. $M_{\text{high harmony}}$ = 6.25, SD = .89), quality (F[1,113] = 9.37; p = .003; η^2 = .08; $M_{\text{low harmony}}$ = 5.18, SD = 1.50 vs. $M_{\text{high harmony}}$ = 5.95, SD = 1.17), customer service (F[1,113] = 10.10; p = .002; η^2 = .08; $M_{\text{low harmony}}$ = 5.67, SD = 1.28 vs. $M_{\text{high harmony}}$ = 6.32, SD = .85), and value for money (F[1,113] = 5.39; p = .022; η^2 = .05; $M_{\text{low harmony}}$ = 5.09, SD = 1.33 vs. $M_{\text{high harmony}}$ = 6.41, SD = 1.10), whereas differences were marginal for atmosphere (F[1,113] = 3.77; p = .055; η^2 = .05; $M_{\text{low harmony}}$ = 6.16, SD = 1.14 vs. $M_{\text{high harmony}}$ = 6.51, SD = .78) and nonsignificant for assortment (F[1,113] = 2.84; p = .095; η^2 = .02; $M_{\text{low harmony}}$ = 6.24, SD = 1.20 vs. $M_{\text{high harmony}}$ = 6.54, SD = .63). Importantly, post hoc power analysis conducted with G*POWER (Faul, Erdfelder, Lang, & Buchner, 2007) revealed acceptable power measures of 0.93 or higher, suggesting that the size of our participant group is sufficient for the analyses generated (Cohen, 1992).

Further, analyses of covariance revealed a significant effect of visual harmony on loyalty (F[1,113] = 3.93; p = .050; η^2 = .08; $M_{\text{low harmony}}$ = 5.74, SD = 1.34 vs. $M_{\text{high harmony}}$ = 6.20, SD = .77) but not on satisfaction (F[1,113] = 2.67; p = .105; η^2 = .04; $M_{\text{low harmony}}$ = 5.89, SD = 1.12 vs. $M_{\text{high harmony}}$ = 6.19, SD = .77). Again, the effect of expertise in completing questionnaires was nonsignificant (p > .10). Taken together, the significant differences in image and loyalty support Hypothesis 1, and the claim that visual harmony in a survey will lead consumers to exhibit more positive scores for constructs assessed on Likert scales.

3.2.2 | Scale reliability

Cronbach's alpha was calculated for the three-item measures of satisfaction and loyalty within each of the two questionnaire designs, using McGraw and Wong's (1996) method to estimate confidence intervals and Hakstian and Whalen's (1976) method to estimate statistical significance of differences in alpha (Diedenhofen & Musch, 2016). For satisfaction, Cronbach's alpha was lower (p = .002) for the high harmony questionnaire (α = .51; LLCI = .24, ULCI = .69; variance explained = .68; item-to-factor loadings = .80, .83, .85) than for the low harmony questionnaire (α = .86; LLCI = .78, ULCI = .91; variance explained = .80; item-to-factor loadings = .84, .91, .93). For loyalty, there was no significant difference between Cronbach's alphas (α = .82 vs. α = .87; p = .393). These findings provide partial support for Hypothesis 2, and the claim that visual harmony in a survey will be detrimental to the reliability of Likert scales.

3.2.3 | Construct validity

Construct validity involves the network of associations between a supposed construct and other constructs predicted by theory. We tested two aspects of construct validity, the theorized and previously found associations between image and satisfaction (e.g., Orth & Green, 2009) and between satisfaction and loyalty (e.g., Bloemer & De Ruyter, 1998).

If the associations varied across the two treatments, it would be evidence against the cross-harmony construct validity of image, satisfaction, or loyalty, or the association of one or several of these constructs with the tools used to measure them. We tested two regressions, the first, testing the impact of image variables on satisfaction and the second, testing the influence of satisfaction on loyalty, using consumer expertise as a covariate in both analyses.[1] With data pooled across conditions, results in Table 1 indicated significant effects of convenience and value for money on satisfaction along with marginal effects of selection (negative) and expertise. With the low harmony, survey satisfaction was influenced by the store's atmosphere, selection, and consumer expertise. In contrast, with the high harmony survey, significant predictors included convenience, quality, value for money, and (marginally) selection. The effect of visual harmony on the associations was significant for four of the six image predictors and was marginal for the other two. Specifically, when directly testing for differences in regression, coefficients between the two harmony conditions, differences were significant for convenience (p = .032), quality (p = .022), customer service (p = .034), and value for money (p = .014) and were marginal for atmosphere (p = .054) and selection (p = .087).

Examining the association between satisfaction and loyalty yielded similar results. Across treatments, satisfaction was a significant predictor of loyalty. The effect of visual harmony on the association between the two constructs was significant (p = .024) with the association being stronger (p = .024) in the low harmony condition (β = .76, p = .001) than in the high harmony condition (β = .28, p = .029). These findings support Hypothesis 3, and the claim that visual harmony in a survey has a negative impact on construct validity.

3.2.4 | Discussion of Study 1 findings

Taken together, Study 1 findings provide initial support for the claim that harmony in a questionnaire's visual design can impair measurement accuracy. Specifically, visually more harmonious surveys lead to greater positivity in several important store image measures, lower reliability of the satisfaction scale, and questionable validity for store image, satisfaction, and loyalty. Although those results are noteworthy and in line with expectations, at least two aspects need further investigating: First, Study 1 findings were obtained with a females-only sample recruited through a mall intercept approach on a single (customer-relationship management) topic, thereby lacking evidence of robustness and generalizability (in addition to lacking data for testing Hypothesis 4). Second, the study varied visual harmony mostly by manipulating specific design elements (colors, spacing, typeface, contrasts, and borders) and to lesser extent relationships between key properties of visual harmony (i.e., coherence, connectedness, symmetry, contrast, and balance), leaving it ambiguous how harmony is created and how findings can be applied by designers. These possible limitations motivated the second study.

3.3 | Study 2

One hundred and eighty[2] members of a consumer panel (mean age of 36.5 years, 51% females) participated in a 2 (harmony: low versus high) × 2 (gender: female versus male) full-factorial experiment involving a soon-to-be marketed e-scooter. In addition to being balanced in terms of gender, subsamples in the "high" versus "low" harmony conditions did not differ in age (p > .10). Two versions of the questionnaire were created by manipulating visual harmony through relational properties (i.e., coherence, connectedness, symmetry, contrast, and balance) of the previously employed elements (see Figure A2). A pretest with members of the target audience (N = 11) established significant differences in visual harmony (F[1,10] = 11.86, p = .007), with the version designed for high harmony scoring higher than the version designed for low

TABLE 1 Testing for differences in regression coefficients (Study 1)

Predictors	Pooled data β (p value)	Low harmony β (p value)	High harmony β (p value)	Difference p value
Outcome: Satisfaction				
Atmosphere	.17 (.100)	.42 (.027)	.03 (.816)	.054
Convenience	.22 (.038)	.10 (.497)	.36 (.011)	.032
Quality	.17 (.102)	.12 (.445)	.29 (.025)	.022
Selection	−.19 (.053)	−.41 (.024)	−.09 (.426)	.087
Customer service	.18 (.123)	.23 (.243)	.20 (.093)	.034
Value for money	.22 (.035)	.10 (.558)	.28 (.022)	.014
Expertise (control)	.16 (085)	.31 (.024)	.09 (.470)	.011
Harmony (manip.)	−.06 (.476)	-	-	-
Outcome: Loyalty				
Satisfaction	.63 (.001)	.76 (.001)	.28 (.029)	.024
Expertise (control)	−.10 (.171)	−.01 (.954)	−.36 (.005)	.091
Harmony (manip.)	.07 (.171)	-	-	-

Note: N = 115.

harmony (M = 5.14, SD = .83 vs. M = 6.63, SD = .63). Appendix B illustrates the stimuli selected for the main study.

In line with the new product and brand introduction context the survey employed previously developed and validated multi-item measures of brand personality (Aaker, 1997), hedonic and utilitarian product benefits (Chitturi, Raghunathan, & Mahajan, 2008), brand attitude (Voss, Spangenberg, & Grohmann, 2003), and purchase intention (Sweeney, Soutar, & Johnson, 1999). In addition, single-item measures assessed consumer perception of elements and key relational properties of visual harmony (Lauer, 1979: e.g., "A congruity or arrangement exists among the elements in the questionnaire's design". "The elements in the questionnaire's design look as though they belong together." "There is some visual connection among the elements in the questionnaire's design that causes them to come together.") Measures of importance of harmony in personal life (IHL; Kwan, Bond, & Singelis, 1997), design acumen (ACU; Bloch, Brunel, & Arnold, 2003), familiarity with questionnaire design (FQD), design naturalness and elaborateness were included as controls ("All in all the visual design of this questionnaire is [natural] [elaborate]"). As with Study 1, all measures were 7-item Likert scales with a left–right descending order, from 7 (fully agree) to 1 (fully disagree).

A manipulation check yielded a significant effect of the treatments on perceived harmony (F[1,179] = 8.38, p = .001), with the stimulus designed for high harmony scoring higher (M = 4.33) than the stimulus designed for low harmony (M = 3.58).

3.3.1 | Pattern of responses

A general linear model was employed to test Hypothesis 1 and Hypothesis 4a with consumer gender and manipulated harmony as the factors, brand personality, utilitarian, and hedonic product benefits, brand attitude and purchase intention as the dependent variables, and IHL, ACU, FQD, design naturalness (NAT) and elaborateness (ELAB) included as covariates. The results (see Table 2) indicate significant effects of the harmony × gender interaction term on several important outcome variables but no main effects of harmony. Specifically, in the presence of several significant effects of control variables, harmony × gender interaction effects were significant (p < .05) for utilitarian benefits (females: $M_{\text{low harmony}}$ = 4.86, SD = 1.35 vs. $M_{\text{high harmony}}$ = 5.17, SD = 1.38; males: $M_{\text{low harmony}}$ = 3.92, SD = 1.39 vs. $M_{\text{high harmony}}$ = 4.94, SD = 1.42) and purchase intention (females: $M_{\text{low harmony}}$ = 2.47, SD = 1.36 vs. $M_{\text{high harmony}}$ = 2.95, SD = 1.43; males: $M_{\text{low harmony}}$ = 2.40, SD = 1.38 vs. $M_{\text{high harmony}}$ = 3.04, SD = 1.39), were marginal (p < .10) for brand personality (females: $M_{\text{low harmony}}$ = 3.72, SD = 1.25 vs. $M_{\text{high harmony}}$ = 3.87, SD = 1.46; males: $M_{\text{low harmony}}$ = 3.42, SD = 1.07 vs. $M_{\text{high harmony}}$ = 4.11, SD = 1.40) and were nonsignificant for brand attitude and hedonic benefits. Post hoc power analysis (Faul et al., 2007) revealed acceptable power measures of 0.94 or higher, suggesting that the size of our participant group is sufficient for the analyses generated (Cohen, 1992). The findings fail to support Hypothesis 1 but partially support Hypothesis 4a, and the claim that visual harmony in a survey will lead female consumers to exhibit more positive scores for constructs assessed on Likert scales.

TABLE 2 Testing for differences in response patterns (Study 2)

Independent variable	Dependent variable	*df*	*F*	*p*	η^2	$M_{low\ HARM}$	$M_{high\ HARM}$
Harmony (HARM)	BP	1	.34	.854	.01	3.64	3.89
	UTI	1	2.47	.118	.02	4.52	4.91
	HED	1	.11	.742	.01	4.45	4.57
	AB	1	.17	.678	.01	4.26	4.31
	PI	1	.37	.552	.01	2.66	2.72
Sex	BP	1	.38	.541	.01	-	-
	UTI	1	7.04	.009	.04	-	-
	HED	1	.39	.531	.01	-	-
	AB	1	2.52	.104	.02	-	-
	PI	1	.29	.593	.01	-	-
HARM*Sex	BP	1	2.64	.106	.02	-	-
	UTI	1	4.49	.036	.03	-	-
	HED	1	.61	.434	.01	-	-
	AB	1	1.59	.209	.01	-	-
	PI	1	4.84	.029	.03	-	-
IHL	BP	1	10.16	.002	.06	-	-
	UTI	1	.34	.563	.01	-	-
	HED	1	1.38	.242	.01	-	-
	AB	1	1.13	.289	.01	-	-
	PI	1	.06	.807	.01	-	-
ACU	BP	1	.07	.789	.01	-	-
	UTI	1	.19	.663	.01	-	-
	HED	1	.01	.917	.01	-	-
	AB	1	1.70	.194	.01	-	-
	PI	1	5.95	.016	.04	-	-
FQD	BP	1	2.25	.136	.01	-	-
	UTI	1	2.03	.156	.01	-	-
	HED	1	2.40	.123	.01	-	-
	AB	1	4.92	.028	.03	-	-
	PI	1	.31	.581	.01	-	-
NAT	BP	1	36.81	.001	.18	-	-
	UTI	1	6.14	.014	.04	-	-
	HED	1	13.29	.001	.07	-	-
	AB	1	9.73	.002	.06	-	-
	PI	1	29.85	.001	.15	-	-
ELAB	BP	1	6.30	.013	.04	-	-
	UTI	1	14.57	.001	.08	-	-
	HED	1	9.99	.002	.06	-	-
	AB	1	5.52	.020	.03	-	-
	PI	1	4.55	.034	.03	-	-

Note: N = 180. IHL: Importance of harmony in personal life.
Abbreviations: AB, attitude towards brand; ACU, design acumen; BP, brand personality; ELAB, design elaborateness; FQD, familiarity with questionnaire design; HED, hedonic product benefits; NAT, design naturalness; PI, purchase intention; UTI, utilitarian product benefits.

3.3.2 | Scale reliability

Cronbach's alpha was calculated for the multi-item measures of brand personality (5 items), hedonic and utilitarian product benefits (3 items, respectively), brand attitude (3 items), and purchase intention (3 items) within each of the two questionnaire designs and the two genders, using the methods previously employed in Study 1 (Diedenhofen & Musch, 2016). When testing for the influence of harmony (across genders), Cronbach's alpha was lower ($p < .05$) for the high harmony questionnaire ($\alpha = .90$; LLCI = .86, ULCI = .93) than for the low harmony questionnaire ($\alpha = .94$; LLCI = .92, ULCI = .96) only for the brand personality measure. All other measures did not exhibit significant differences in reliability between harmony conditions. When additionally accounting for the influence of gender, Cronbach's alpha was lower for the high harmony questionnaire with females (compared with males) for brand attitude (females: $\alpha = .91$; LLCI = .85, ULCI = .95; males: $\alpha = .97$; LLCI = .95, ULCI = .98, $p = .015$), hedonic benefits (females: $\alpha = .87$; LLCI = .78, ULCI = .92; males: $\alpha = .95$; LLCI = .92, ULCI = .97, $p = .018$), and brand personality (females: $\alpha = .87$; LLCI = .79, ULCI = .91; males: $\alpha = .94$; LLCI = .92, ULCI = .97, $p = .047$). Differences for purchase intention and utilitarian benefits were nonsignificant. These findings partially support Hypothesis 2, and the claim that visual harmony in a survey will be detrimental to the reliability of Likert scales. Similarly, Hypothesis 4b and the claim that consumer gender will interact with harmony to impact reliability is partially supported.

3.3.3 | Construct validity

To examine harmony's impact on the network of associations between constructs, we employed conditional process modeling (Hayes, 2013; Model #72). Specifically, we tested a moderated mediation model where harmony moderates the relationships between brand personality and brand attitude, between product benefits and brand attitude, and between brand attitude and purchase intention, with each moderating effect of harmony being moderated by consumer gender. IHL, ACU, FQD, design naturalness, and elaborateness were included as covariates.

Results indicate that brand attitude is significantly influenced by utilitarian benefits ($B = .64$, $SE = .07$, $p = .001$) and hedonic benefits ($B = .35$, $SE = .07$, $p = .001$), as well as the utilitarian × harmony interaction term ($B = .31$, $SE = .15$, $p = .038$). Purchase intention is influenced by brand personality ($B = .29$, $SE = .07$, $p = .001$), hedonic benefits ($B = .16$, $SE = .08$, $p = .020$), and design naturalness ($B = .14$, $SE = .06$, $p = .017$). Most important, bootstrap results indicate several differences in the conditional indirect effects. Specifically, the effect of utilitarian benefits, through brand attitude, on purchase intention in the high harmony condition was significant and strong with females ($B = .27$, $SE = .10$, LLCI = .06, ULCI = .47), whereas it was nonsignificant with males (LLCI = −.01, ULCI = .37). Effects in the low harmony condition were significant for both females ($B = .15$, $SE = .08$, LLCI = .01, ULCI = .32) and males ($B = .17$, $SE = .09$, LLCI = .01, ULCI = .35). Similarly, the effect of hedonic benefits (through brand attitude) on purchase intention in the high harmony condition was significant with females ($B = .16$, $SE = .08$, LLCI = .03, ULCI = .36) but not with males (LLCI = −.01, ULCI = .19). In the low harmony condition, indirect hedonic effects were significant with males ($B = .10$, $SE = .06$, LLCI = .01, ULCI = .22) but not with females (LLCI = −.02, ULCI = .21). All other effects were nonsignificant. Together, these findings partially support Hypothesis 3 and Hypothesis 4c, especially the claim that visual harmony in a survey has a negative impact on construct validity contingent upon respondent gender.

3.3.4 | Drivers of overall visual harmony

Finally, to provide practitioners with direction on how to calibrate visual harmony in a questionnaire's design, we employed correlation analyses to examine (a) relationships between an overall measure of perceived harmony and relational properties and (b) relationships between overall harmony and specific design elements. Regarding relational properties, the results indicate that the harmony consumers perceive in a questionnaire's design correlates significantly and positively with balance ($r = .65$, $p < .001$), coherence ($r = .65$, $p < .001$), connectedness ($r = .54$, $p < .001$), and symmetry ($r = .54$, $p < .001$), and to a lesser extent with contrast ($r = .29$, $p < .001$). Regarding elements, overall harmony correlated significantly and positively with harmony in text boxes ($r = .48$, $p < .001$), alignments ($r = .50$, $p < .001$), spacing ($r = .46$, $p < .001$), shapes ($r = .55$, $p < .001$), colors ($r = .48$, $p < .001$), type font ($r = .48$, $p < .001$), and images ($r = .60$, $p < .001$). These findings highlight the potential of relational properties and specific design elements for creating desired levels of overall visual harmony in a questionnaire's design.

4 | GENERAL DISCUSSION

In two studies, visual harmony in a survey instrument affected consumer response in several ways including pattern of responses, scale reliability, and construct validity. The effect of visual harmony may stem from visual priming, which increases accessibility of the concept and makes agreement a more focal concept, thereby shifting responses to Likert scales toward the positive side (i.e., greater agreement).

If a survey is used to make relative judgments between customers or groups of customers (e.g., segments) on the basis of overall scores, the effect of visual harmony is not critical because all respondents are affected by the biases. Caution needs to be exercised, however, when absolute judgments are made. For example, using the data collected in Study 1, if the provider was to set a performance criterion at a service quality score of 6.0, it would be met by responses to the high harmony design, but not by responses to the low harmony design. In addition, although overall satisfaction scores (and two of the image constructs) did not show an effect of visual harmony, large differences were evident when downstream effects were considered. For

example, using data collected with the low harmony questionnaire, one might conclude that increasing satisfaction by one unit will lead to an increase in loyalty by .76 units, implying that investing into satisfaction should be a worthwhile endeavor. We have shown, however, that visual harmony in a survey can significantly inflate associations between constructs, as the corresponding coefficient for data collected with the high harmony questionnaire is a mere .28. Similarly, if a service provider's image campaign is contingent upon survey results identifying specific facets (i.e., atmosphere, convenience, quality, selection, customer service, and value for money), the relative impact and significance of those drivers will change significantly depending on the visual harmony of the survey instrument. Visual harmony in a measurement instrument may even impact scale reliability. Although our finding of differences in alphas are ambiguous as they occurred with a few, but not all scales, and especially with females, this finding is rather disturbing and warrants further investigation. If the pattern of differences reported here were replicated in other scales, possibly including more items, it would imply that the meaning of the Likert format could change depending on the visual design of the survey.

Overall, this study provides evidence suggesting that visual harmony in a survey can detrimentally impact respondents' behavior, in line with recent reports that appearance alone can potentially impact data quality (Deutskens, de Ruyter, Wetzels, & Oosterveld, 2004; Mahon-Haft & Dillman, 2010). If reactions to a compact and fairly simple paper and pencil survey with a design high versus low in visual harmony can have such an impact among a customer population, how might reactions vary in other contexts such as more visually elaborate online surveys or on more personal individual variables? Therefore, even the relatively minor harm to measurement accuracy observed here suggests that service providers and researchers need to carefully calibrate visual harmony when designing surveys.

Unfortunately, our research does not pinpoint contexts or details where harmony impairs measurement accuracy. Instead, in both a customer relationship management and a new product and brand introduction context, a survey's visual harmony was detrimental to some response patterns and reliabilities but not to others. Perhaps the most obvious implication of our findings thus is to alert practitioners to the possibility that visual harmony has the potential to bias response patterns, affect reliabilities of measures, and impair construct validity. Across contexts, surveys are frequently designed to be visually appealing, thus adopting design principles related to harmony (Haberstroh et al., 2018). Our findings suggest that high levels of visual harmony in a design should be avoided, especially in combination with Likert-type scales and when researching female consumers. On the other hand, low harmony levels should be avoided as well as they similarly can fail to yield realistic and valid results.

To achieve intermediate levels of visual harmony managers can draw from a variety of elements and relational properties. For example, according to Study 2 findings, overall visual harmony traces back to specific elements such as text boxes, type font, shapes, images, and alignment. In a first step, each of those elements should be considered to ascertain that it meets the desired level of harmony. For example, more and less harmonious type fonts can be selected following Henderson et al. (2004) guidelines on harmony in colors, and contrasts are available from Schloss and Palmer (2011) and Tourangeau et al. (2004) aid in designing borders. In a second step, relational properties between those elements should be considered to ascertain appropriate levels of balance, coherence, connectedness, symmetry, and contrast. In a final step, researchers may wish to include an overall measure of visual harmony in pretesting the measurement instrument to better calibrate their design.

The finding that harmony influences consumer responses when a number of control variables are accounted for also has implications for managerial practice. Specifically, harmony exerts its influence regardless of the expertise consumers have with completing surveys, their familiarity with the visual design of questionnaires, and in the presence of significant effects of the design's naturalness and elaborateness. Together, these findings highlight the robustness of harmony effects, indicating that they occur not only under a narrow set of highly specific circumstances but represent a broader phenomenon worthy of managerial attention.

There are several limitations to this research. First, harmony, our study's focal concept, is one among several factors of visual design (Henderson et al., 2004). Yet extant literature is scarce on how visual harmony impacts viewer response relative to other factors (Orth, Campana, & Malkewitz, 2010). Given Study 2's finding of significant effects of visual harmony in the presence of effects of other design factors (i.e., naturalness and elaborateness), designers can be more confident that paying attention to visual harmony is a worthwhile endeavor. We speculate that the relative importance of harmony may depend on the context. For example, harmony effects may become weaker when factors such as visual complexity hinder processing by depleting cognitive resources (Orth & Crouch, 2014). Conversely, the influence of visual harmony may increase when congruent input is received through other sensory modalities, such as haptics (Littel & Orth, 2013). Future research may find it worthwhile to investigate these and other boundary conditions.

Second, our findings do not uniformly associate visual harmony with differences in response patterns, reliabilities, and construct validities. For example, differences in Study 1 response patterns emerge for satisfaction but not for loyalty. In Study 2, differences emerge with females in purchase intention and utilitarian but not hedonic product benefits. Although this may not be surprising given the more embryonic nature of our approach, researchers may find it fruitful to further explore the reasons underlying divergent effects.

ORCID

Ulrich R. Orth https://orcid.org/0000-0003-1134-9892

ENDNOTES

[1] Readers should note that, although the small Cronbach's alpha of the satisfaction measure in the high harmony condition constitutes a possible limitation, unbundling both the test and the sample in our tests for construct validity is statistically sound (Bernardi, 1994).

[2] The issue of a priori calculation of the test design (i.e., computing the appropriate sample size) was addressed by adopting the procedure

prescribed by Faul, Erdfelder, Buchner, and Lang (2009) and following Lakens (2013). Given the limited time available for collecting Study 2 data, the actual sample size (N = 180) is slightly smaller than the target value of 200 respondents. Because the time required for data collection was critical, this sample size can be considered acceptable (Faul et al., 2007).

REFERENCES

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research*, *34*(3), 347–356. https://doi.org/10.2307/3151897

Aspara, J., & Van Den Bergh, B. (2014). Naturally designed for masculinity vs. femininity? Prenatal testosterone predicts male consumers' choices of gender-imaged products. *International Journal of Research in Marketing*, *31*(1), 117–121. https://doi.org/10.1016/j.ijresmar.2013.09.001

Bar, M., & Biederman, I. (1998). Subliminal visual priming. *Psychological Science*, *9*, 464–468. https://doi.org/10.1111/1467-9280.00086

Bernardi, R. A. (1994). Validating research results when Cronbach's alpha is below. 70: A methodological procedure. *Educational and Psychological Measurement*, *54*(3), 766–775. https://doi.org/10.1177/0013164494054003023

Bloch, P. H., Brunel, F. F., & Arnold, T. J. (2003). Individual differences in the centrality of visual product aesthetics: Concept and measurement. *Journal of Consumer Research*, *29*(4), 551–565. https://doi.org/10.1086/346250

Bloemer, J., & De Ruyter, K. (1998). On the relationship between store image, store satisfaction and store loyalty. *European Journal of Marketing*, *32*, 499–513. https://doi.org/10.1108/03090569810216118

Bruner, G. C. (2015). *Multi-item measures for marketing and consumer behavior research*, *Handbook of Marketing Scales*. Chicago, Illinois: Sage Publications, Vol. 7.

Chamorro-Premuzic, T., Burke, C., Hsu, A., & Swami, V. (2010). Personality predictors of artistic preferences as a function of the emotional valence and perceived complexity of paintings. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 196–204. https://doi.org/10.1037/a0019211

Chang, D., & Nesbitt, K. V. (2006, April). Developing Gestalt-based design guidelines for multi-sensory displays. In Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop-Volume 57 (pp. 9-16). Australian Computer Society, Inc.

Chatterjee, A. (2010). Neuroaesthetics: A coming of age story. *Journal of Cognitive Neuroscience*, *23*(1), 53–62. https://doi.org/10.1162/jocn.2010.21457

Chitturi, R., Raghunathan, R., & Mahajan, V. (2008). Delight by design: The role of hedonic versus utilitarian benefits. *Journal of Marketing*, *72*(3), 48–63. https://doi.org/10.1509/JMKG.72.3.048

Chowdhury, J., Reardon, J., & Srivastava, R. (1998). Alternative modes of measuring store image: An empirical assessment of structured versus unstructured measures. *Journal of Marketing Theory and Practice*, *6*, 72–86. https://doi.org/10.1080/10696679.1998.11501797

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping respondents get it right the first time: The influence of words, symbols, and graphics in web surveys. *Public Opinion Quarterly*, *71*, 113–125. https://doi.org/10.1093/poq/nfl039

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159: https://doi.org/10.1037//0033-2909.112.1.155

Crilly, N., Moultrie, J., & Clarkson, P. J. (2004). Seeing things: Consumer response to the visual domain in product design. *Design Studies*, *25*, 547–577. https://doi.org/10.1016/j.destud.2004.03.001

Darley, W. K., & Smith, R. E. (1995). Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. *Journal of Advertising*, *24*(1), 41–56. https://doi.org/10.1080/00913367.1995.10673467

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, *15*, 21–36. https://doi.org/10.1023/B:MARK.0000021968.86465.00

Diedenhofen, B., & Musch, J. (2016). COCRON: A web interface and R package for the statistical comparison of Cronbach's alpha coefficients. *International Journal of Internet Science*, *11*, 51–60. https://www.ijis.net/ijis11_1/ijis11_1_diedenhofen_and_musch.html

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Haberstroh, K., Orth, U. R., Bouzdine-Chameeva, T., Cohen, J., Corsi, A. M., Crouch, R., & De Marchi, R. (2018). Through the lens of self-construal: Cross-cultural variation in consumers' appreciation of harmony in marketing visuals. *International Marketing Review*, *35*, 429–457. https://doi.org/10.1108/IMR-12-2015-0283

Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, *41*, 219–231. https://doi.org/10.1007/BF02291840

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis—A regression-based approach. *New York, NY: Guilford Press*, *51*, 335–337. https://doi.org/10.1111/jedm.12050

Hekkert, P. (2006). Design aesthetics: Principles of pleasure in design. *Psychological Science*, *48*(2), 157–172.

Henderson, P. W., & Cote, J. A. (1998). Guidelines for selecting or modifying logos. *Journal of Marketing*, *62*, 14–30. https://doi.org/10.2307/1252158

Henderson, P. W., Giese, J. L., & Cote, J. A. (2004). Impression management using typeface design. *Journal of Marketing*, *68*, 60–72. https://doi.org/10.1509/jmkg.68.4.60.42736

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, *19*, 323–340. https://doi.org/10.1037/0022-3514.51.6.1106

Jiang, Y., Gorn, G. J., Galli, M., & Chattopadhyay, A. (2016). Does your company have the right logo? How and why circular- and angular-logo shapes influence brand attribute judgments. *Journal of Consumer Research*, *42*, 709–726. https://doi.org/10.1093/jcr/ucv049

Koffka, K. (1935). *Principles of Gestalt psychology*. New York, NY: Harcourt, Brace and World.

Kumar, M., & Garg, N. (2010). Aesthetic principles and cognitive emotion appraisals: How much of the beauty lies in the eye of the beholder. *Journal of Consumer Psychology*, *20*, 485–494. https://doi.org/10.1016/j.jcps.2010.06.015

Kumar, M., & Noble, C. H. (2016). Beyond form and function: Why do consumers value product design? *Journal of Business Research*, *69*, 613–620. https://doi.org/10.1016/j.jbusres.2015.05.017

Kwan, V. S., Bond, M. H., & Singelis, T. M. (1997). Pancultural explanations for life satisfaction: Adding relationship harmony to self-esteem. *Journal of Personality and Social Psychology*, *73*(5), 1038–1051: https://doi.org/10.1037//0022-3514.73.5.1038

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lauer, D. (1979). *Design basics*. New York: Holt, Rinehart and Winston.

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing & Health*, *25*, 295–306. https://doi.org/10.1002/nur.10041

Likert, R. (1932). A technique for the measurement of attitudes. In R. S. Woodworth (Ed.), *Archives of Psychology*. (Vol. 22, pp. 5–55) New York, NY: Columbia University Press.

Littel, S., & Orth, U. R. (2013). Effects of package visuals and haptics on brand evaluations. *European Journal of Marketing*, *47*(1/2), 198–217. https://doi.org/10.1108/03090561311285510

Mahon-Haft, T. A., & Dillman, D. A. (2010). Does visual appeal matter? Effects of web survey aesthetics on survey quality. *Survey Research Methods*, *4*, 43–59. https://doi.org/10.18148/srm/2010.v4i1.2264

Mandel, N., & Johnson, E. J. (2002). When web pages influence choice: Effects of visual primes on experts and novices. *Journal of Consumer Research*, *29*, 235–245. https://doi.org/10.1086/341573

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Meyers-Levy, J. (1989). Gender differences in information processing: A selective interpretation. In P. Cafferata & A. Tybout (Eds.), *Cognitive and affective responses to advertising* (pp. 219–260). Lexington, MA: Lexington Books.

Meyers-Levy, J., & Loken, B. (2015). Revisiting gender differences: What we know and what lies ahead. *Journal of Consumer Psychology*, *25*(1), 129–149. https://doi.org/10.1016/j.jcps.2014.06.003

Minton, E. A., Cornwell, T. B., & Kahle, L. R. (2017). A theoretical review of consumer priming: Prospective theory, retrospective theory, and the affective–behavioral–cognitive model. *Journal of Consumer Behaviour*, *16*, 309–321. https://doi.org/10.1002/cb.1624

Moss, G. (2009). *Gender, design and marketing: How gender drives our perception of design and marketing*. New York, NY: Gower Publishing Ltd. https://doi.org/10.4324/9781315254593

Nicholls, M. E. R., Orr, C. A., Okubo, M., & Loftus, A. (2006). Satisfaction guaranteed: The effect of spatial biases on responses to Likert scales. *Psychological Science*, *17*, 1027–1028. https://dx.doi.org/10.1111/j.1467-9280.2006.01822.x

Noble, C. H., & Kumar, M. (2010). Exploring the appeal of product design: A grounded, value-based model of key design elements and relationships. *Journal of Product Innovation Management*, *27*(5), 640–657. https://doi.org/10.1111/j.1540-5885.2010.00742.x

Orth, U. R., Campana, D., & Malkewitz, K. (2010). Formation of consumer price expectation based on package design: Attractive and quality routes. *Journal of Marketing Theory and Practice*, *18*(1), 23–40. https://doi.org/10.2753/MTP1069-6679180102

Orth, U. R., & Crouch, R. C. (2014). Is beauty in the aisles of the retailer? Package processing in visually complex contexts. *Journal of Retailing*, *90*(4), 524–537. https://doi.org/10.1016/j.jretai.2014.05.004

Orth, U. R., & Green, M. T. (2009). Consumer loyalty to family versus non-family business: The roles of store image, trust and satisfaction. *Journal of Retailing and Consumer Services*, *16*, 248–259. https://doi.org/10.1016/j.jretconser.2008.12.002

Orth, U. R., & Malkewitz, K. (2008). Holistic package design and consumer brand impressions. *Journal of Marketing*, *72*, 64–81. https://doi.org/10.1509/jmkg.72.3.64

Pittard, N., Ewing, M., & Jevons, C. (2007). Aesthetic theory and logo design: Examining consumer response to proportion across cultures. *International Marketing Review*, *24*, 457–473. https://doi.org/10.1108/02651330710761026

Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, *6*(6–7), 15–51.

Schloss, K. B., & Palmer, S. E. (2011). Aesthetic response to color combinations: Preference, harmony, and similarity. *Attention, Perception, & Psychophysics*, *73*, 551–571. https://doi.org/10.3758%2Fs13414-010-0027-0

Singer, E. (2018). The use and effects of incentives in surveys. In *The Palgrave Handbook of Survey Research* (pp. 63–70). Champions: Palgrave Macmillan.

Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, *66*, 15–37. https://doi.org/10.1509/jmkg.66.1.15.18449

Smith-Gratto, K., & Fisher, M. M. (1999). Gestalt theory: A foundation for instructional screen design. *Journal of Educational Technology Systems*, *27*(4), 361–371. https://doi.org/10.2190/KVVE-B0EL-B0CJ-92KM

Sweeney, J. C., Soutar, G. N., & Johnson, L. W. (1999). The role of perceived risk in the quality-value relationship: A study in a retail environment. *Journal of Retailing*, *75*(1), 77–105. https://doi.org/10.1016/S0022-4359(99)80005-0

Toepoel, V., & Funke, F. (2018). Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Mathematical Population Studies*, *25*, 112–122. https://doi.org/10.1080/08898480.2018.1439245

Tourangeau, R. (2018). Choosing a mode of survey data collection. In *The Palgrave Handbook of Survey Research* (pp. 43–50). Champions: Palgrave Macmillan.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*, 368–393. https://doi.org/10.1093/poq/nfh035

Van der Lans, R., Cote, J. A., Cole, C. A., Leong, S. M., Smidts, A., Henderson, P. W., Bluemelhuber, C, Bottomley, P. A., Doyle, J. R, Fedorikhin, A., Moorthy, J., Ramaseshan, B., & Schmitt, B. H. (2009). Cross-national logo evaluation analysis: An individual-level approach. *Marketing Science*, *28*, 968–985: https://doi.org/10.1287/mksc.1080.0462

Voss, K. E., Spangenberg, E. R., & Grohmann, B. (2003). Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of Marketing Research*, *40*(3), 310–320. https://doi.org/10.1509/jmkr.40.3.310.19238

Xue, L., & Yen, C. C. (2007). Towards female preferences in design—A pilot study. *International Journal of Design*, *1*(3), 11–27.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*, 31–46. https://doi.org/10.2307/1251929

AUTHOR BIOGRAPHIES

Ulrich Orth is Professor of Marketing and Chair of A&F Marketing—Consumer Psychology at Kiel University (CAU), Germany. He earned both his doctorate and super-PhD. (Habilitation) from Munich University of Technology, Germany. Past positions include Professor at Mendel University, Czech Republic, and Associate Professor at Oregon State University. Research interests center on consumer behavior and psychology related to visual design with articles published in the Journal of Service Research, Journal of Marketing, Journal of Retailing, Journal of Business Research, Journal of Advertising, Journal of Social Psychology, International Journal of Research in Marketing, among others.

Kristina Nickel is a Post-Doctoral Researcher with the A&F Marketing—Consumer Psychology chair at Kiel University (CAU), Germany. She holds a doctorate from CAU, and her research interests are on visual communication, gender marketing, and consumer behavior.

Rebekka Böhm is a PhD candidate and Graduate Researcher with the A&F Marketing—Consumer Psychology chair at Kiel

University (CAU), Germany. She holds, ora M.Sc. from Maastricht University and is interested in researching survey methods and consumer behavior.

Katharina Röwe is a Graduate Researcher with the A&F Marketing—Consumer Psychology chair at Kiel University (CAU), Germany. She holds a M.Sc. from CAU and is interested in researching survey methods and consumer behavior.

How to cite this article: Orth UR, Nickel K, Böhm R, Röwe K. Agreement by design: The effect of visual harmony on responses to surveys. *J Consumer Behav*. 2020;19:196–207. https://doi.org/10.1002/cb.1804

APPENDIX A.

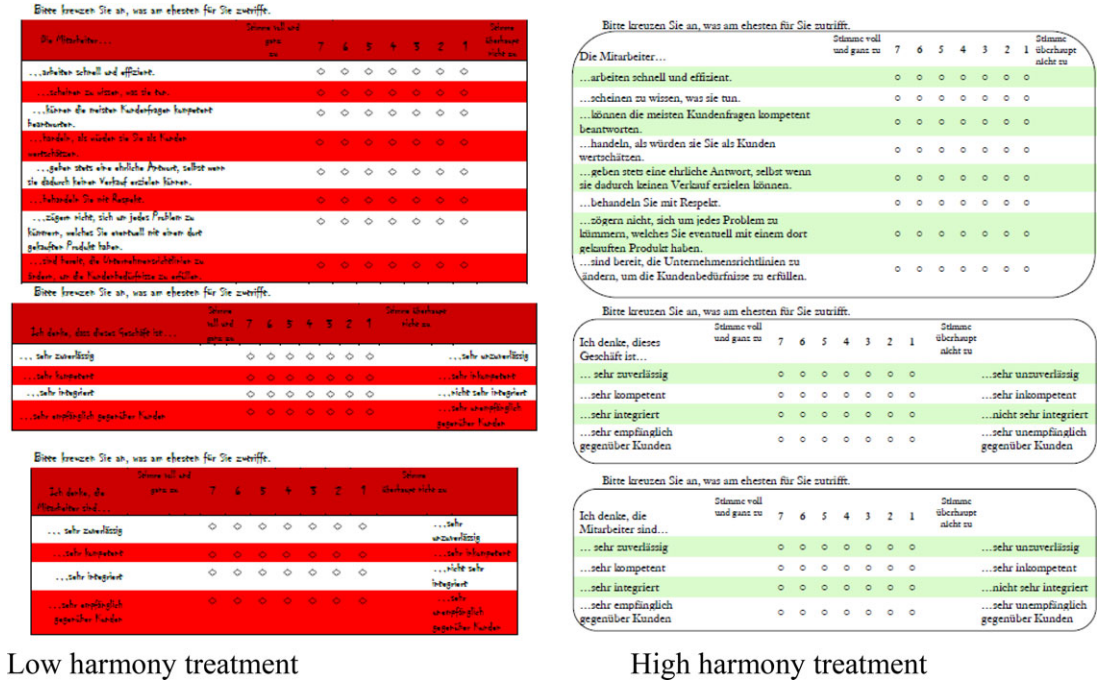
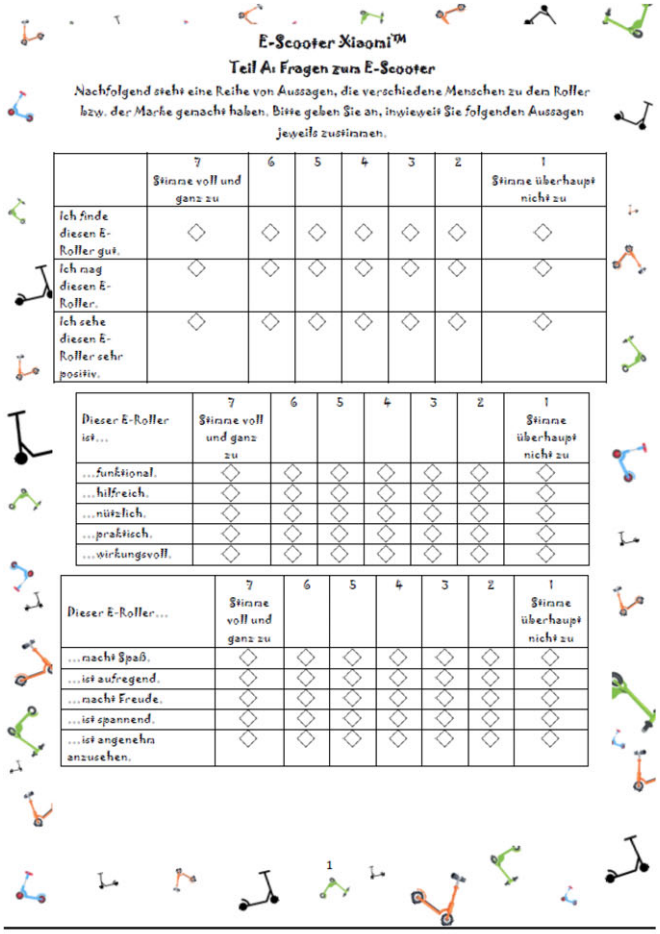


FIGURE A1 Study 1 experimental treatments (one-page excerpts) [Colour figure can be viewed at wileyonlinelibrary.com]



E-Scooter Xiaomi™

Teil A: Fragen zum E-Scooter

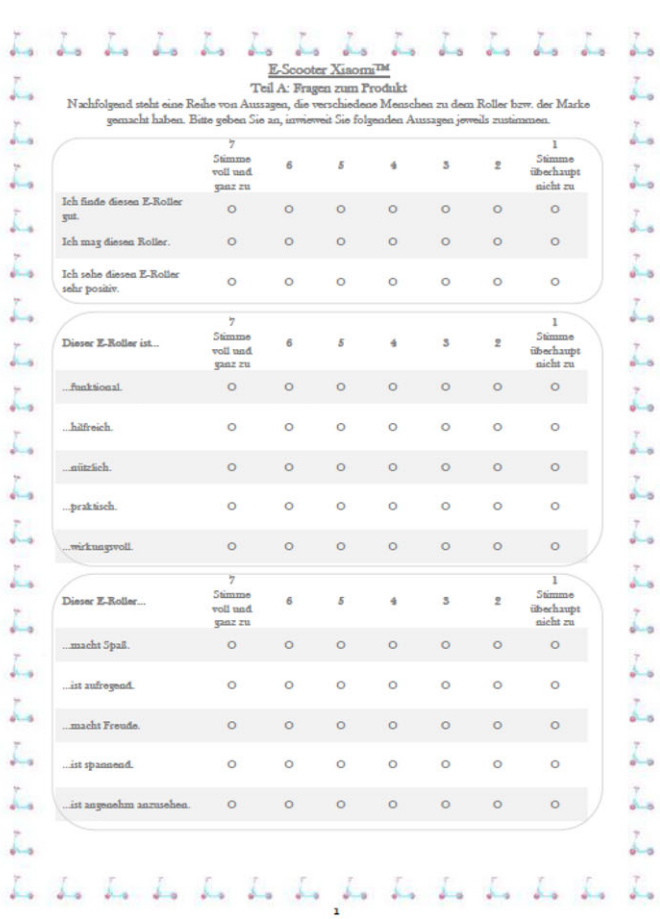
Nachfolgend steht eine Reihe von Aussagen, die verschiedene Menschen zu dem Roller bzw. der Marke gemacht haben. Bitte geben Sie an, inwieweit Sie folgenden Aussagen jeweils zustimmen.

	7 Stimme voll und ganz zu	6	5	4	3	2	1 Stimme überhaupt nicht zu
Ich finde diesen E-Roller gut.	◇	◇	◇	◇	◇	◇	◇
Ich mag diesen E-Roller.	◇	◇	◇	◇	◇	◇	◇
Ich sehe diesen E-Roller sehr positiv.	◇	◇	◇	◇	◇	◇	◇

Dieser E-Roller ist…	7 Stimme voll und ganz zu	6	5	4	3	2	1 Stimme überhaupt nicht zu
…funktional.	◇	◇	◇	◇	◇	◇	◇
…hilfreich.	◇	◇	◇	◇	◇	◇	◇
…nützlich.	◇	◇	◇	◇	◇	◇	◇
…praktisch.	◇	◇	◇	◇	◇	◇	◇
…wirkungsvoll.	◇	◇	◇	◇	◇	◇	◇

Dieser E-Roller…	7 Stimme voll und ganz zu	6	5	4	3	2	1 Stimme überhaupt nicht zu
…macht Spaß.	◇	◇	◇	◇	◇	◇	◇
…ist aufregend.	◇	◇	◇	◇	◇	◇	◇
…macht Freude.	◇	◇	◇	◇	◇	◇	◇
…ist spannend.	◇	◇	◇	◇	◇	◇	◇
…ist angenehm anzusehen.	◇	◇	◇	◇	◇	◇	◇

1



E-Scooter Xiaomi™

Teil A: Fragen zum Produkt

Nachfolgend steht eine Reihe von Aussagen, die verschiedene Menschen zu dem Roller bzw. der Marke gemacht haben. Bitte geben Sie an, inwieweit Sie folgenden Aussagen jeweils zustimmen.

	7 Stimme voll und ganz zu	6	5	4	3	2	1 Stimme überhaupt nicht zu
Ich finde diesen E-Roller gut.	○	○	○	○	○	○	○
Ich mag diesen Roller.	○	○	○	○	○	○	○
Ich sehe diesen E-Roller sehr positiv.	○	○	○	○	○	○	○

Dieser E-Roller ist…	7 Stimme voll und ganz zu	6	5	4	3	2	1 Stimme überhaupt nicht zu
…funktional.	○	○	○	○	○	○	○
…hilfreich.	○	○	○	○	○	○	○
…nützlich.	○	○	○	○	○	○	○
…praktisch.	○	○	○	○	○	○	○
…wirkungsvoll.	○	○	○	○	○	○	○

Dieser E-Roller…	7 Stimme voll und ganz zu	6	5	4	3	2	1 Stimme überhaupt nicht zu
…macht Spaß.	○	○	○	○	○	○	○
…ist aufregend.	○	○	○	○	○	○	○
…macht Freude.	○	○	○	○	○	○	○
…ist spannend.	○	○	○	○	○	○	○
…ist angenehm anzusehen.	○	○	○	○	○	○	○

1

Low harmony treatment

High harmony treatment

FIGURE A2 Study 2 experimental treatments (one-page excerpts) [Colour figure can be viewed at wileyonlinelibrary.com]